DOCUMENT RESUME

ED 091 770 CS 201 375

AUTHOR Beardsmore, H. Baetens; Renkin, A.
TITLE A Test of Spoken English.
PUB DATE 71
NOTE 16p.; Reprint fror "International Review of Applied Linsguistics in Language Teaching," Volume 9, 1971; See related documents CS 201 320-375

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Adult Education; *Diagnostic Tests; *Educational Research; *English (Second Language); *Measurement Instruments; Post Secondary Education; Research Tools; Resource Materials; *Standard Spoken Usage
IDENTIFIERS *The Research Instruments Project; TRIP

ABSTRACT

Designed to test the degree of accuracy, fluency, and intelligibility of the subjects in a manner which will produce information similar to that which could be obtained by an interview technique, the test is based on the idea of fictitious dialogue. The test items were selected in order to leave little scope for variations in the answers so as to establish the individual's ability to manipulate spoken language in an active way. The degree of difficulty and discriminatory power of each item was estimated and expressed on a linear scale, and the distribution of total scores was normal with a mean of 14.96 and a standard deviation of 5.89. [This document is one of those reviewed in The Research Instruments Project (TRIP) monograph "Measures for Research and Evaluation in the English Language Arts" to be published by the Committee on Research of the National Council of Teachers of English in cooperation with the ERIC Clearinghouse on Reading and Communication Skills. A TRIP review which precedes the document lists its category (Standard English as a Second Language or Dialect), title, authors, date, and age range (postsecondary, adult), and describes the instrument's purpose and physical characteristics.] (RB)

ED 091770

NCTE Committee on Research

The Research Instruments Project (TRIP)

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The attached document contains one of the measures reviewed in the TRIP committee monograph titled:

Measures for Research and Evaluation
in the English Language Arts

TRIP is an acronym which signifies an effort to abstract and make readily available measures for research and evaluation in the English language arts. These measures relate to language development, listening, literature, reading, standard English as a second language or dialect, teacher competencies, or writing. In order to make these instruments more readily available, the ERIC Clearinghouse on Reading and Communication Skills has supported the TRIP committee sponsored by the Committee on Research of the National Council of Teachers of English and has processed the material into the ERIC system. The ERIC Clearinghouse accession numbers that encompass most of these documents are CS 201 320-CS 201 375.

TRIP Committee:

W.T. Fagan, Chairman
University of Alberta, Edmonton

Charles R. Cooper
State University of New York
at Buffalo

Julie M. Jensen
The University of Texas at Austin

Bernard O'Donnell
Director, ERIC/RCS

Roy C. O'Donnell
The University of Georgia
Liaison to NCTE Committee
on Research

CS 201 375

NATIONAL COUNCIL OF TEACHERS OF ENGLISH
1111 KENYON ROAD
URBANA, ILLINOIS 61801

International Review of Applied

Linguistics in Language Teaching

Volume 9, No. 1, pp. 1-11

IRAL © Copyright, 1971

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

IRAL

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Category: Standard English as a Second Language or Dialect

Title: "A Test of Spoken English"

Authors: H. Baetens Beardsmore and A. Renkin

Age Range: Post secondary - Adult

Description of Instrument:

Purpose - To test the degree of accuracy, fluency and intelligibility in a manner which will produce information similar to that which could be obtained by an interview technique, while at the same time being more objective and less time-consuming.

Date of Construction - 1971

Physical Description - The test is used at the Institute of Phonetics of the University of Brussels in order to determine a student's level of proficiency for placement in audio-oral English language courses. The aim of instruction is an active and practical working knowledge of English, not the production of language specialists. Because intelligibility is the goal, minor defects of pronunciation, accentuation, and intonation are ignored insofar as they do not alter the meaning of the communication.

The test is based upon the idea of fictitious dialogue. The student reacts to specific and grammatically correct stimuli which determine the response to be provided. He participates in the dialogue with a tape recorded voice so as to establish his ability to manipulate spoken language in an active way that simulates interview behavior. The test items were selected in order to leave little scope for variations in the answers. Items were tried on a variety of native speakers including young children. Item content was based on contrastive analyses of the two languages involved, and on what was considered basic to an elementary knowledge of spoken English. Questions were avoided which brought into play the level of intelligence or general knowledge of the student. Before each section of items, instructions are given in the native language. Examples of English responses are then provided. The following are illustrations of the basic abilities tested:

1) The manipulation of numbers and dates e.g.
   S: Tell me when you were born?
   R: I was born on the fifth of January, 1950.
2) The correct use of adverbs, e.g.
   S: What do you usually drink at breakfast?
   R: I usually/normally drink tea at breakfast.

3) The ability to provide simple personal factual information, e.g.
   S: How did you come here today?
   R: I came by tram/on foot/in my father's car.
4) The ability to agree with a speaker using the basic tenses of regular and irregular verbs (this usage of the verb was a criterion employed throughout the test), e.g.
   S: I liked that novel, did you?
   R: Yes, I did.
5) The ability to express disagreement with a speaker, e.g.
   S: Will you be in on Sunday?
   R: No, I won't.
6) The ability to put a question to a speaker, e.g.
   S: Ask me if I've been to Spain.
   R: Have you been to Spain?
7) The ability to indicate comprehension of a sequence of events or the relationship between cause and effect, e.g.
   S: She bought a ticket, then she caught the train; what did she do before she caught the train?
   R: She bought a ticket.
8) The use of the simple past tense of irregular verbs and the natural use of personal pronouns in dialogue, e.g.
   S: Did you meet my sister?
   R: Yes, I met your sister yesterday.
9) The correct use of the conditional, e.g.
   S: She won't do any work because she's tired. And if she wasn't tired?
   R: She'd do some work.
10) The correct use of the perfect conditional, e.g.
   S: She wouldn't have seen him because he didn't arrive on time; and if he'd arrived on time?
   R: She'd have seen him.

A printed correction sheet is available which provides the most probable answers. The criterion to be adhered to in evaluating responses is intelligibility to the native speaker. Administration requires seventeen minutes, scoring approximately five minutes.

Validity, Reliability and Normative Data:

Available statistical data refer to the first thirty test items, and not to the two final test sections consisting of ten items. The degree of difficulty and discriminatory power of each item was estimated and expressed as a mark varying from 0 to 100 on a linear scale. An average item difficulty of 50.77 and an average discrimination of 38.00 was calculated for 180 French-speaking subjects. Difficulty range was 47, or 72 to 25 on the linear scale. Discrimination range was 46, or 67 to 21 on the scale. The test is interpreted, therefore, to be of moderate difficulty, with no extremely easy or difficult items. The distribution of total scores was normal, with a mean of 14.96 and a standard deviation of 5.89. A Spearman-Brown, split-half reliability coefficient was calculated at .86. Because this test is never used in isolation, but as

part of a battery of five measures, reliability is said to be enhanced. Given no satisfactory external criterion, validity has not yet been estimated.

Ordering Information:

EDRS

Related Documents:

Beardsmore, H. B. and Renkin, A. A test of spoken English. International Review of Applied Linguistics in Language Teaching, 1971, 9, 1-11. (EJ 042 870)

from
Librairie Marcel Didier
4 et 6 rue de la Sorbonne, Paris V

from
Oxford University Press
Press Road Neasden London N.W. 10/England

from Oversea Offices in:
Toronto, Melbourne, Wellington, Bombay, Calcutta, Madras, Karachi, Lahore, Dacca, Cape Town, Johannesburg, Salisbury, Ibadan, Apapa, Nairobi, Lusaka, Addis Ababa, Kuala Lumpur, Tokyo, Hong Kong

from
The Center for Curriculum Development, Inc.
A Chilton Subsidiary
401 Walnut Street
Philadelphia, Pa. 19106

from
Julius Groos Verlag
P. O. Box 629
6900 Heidelberg/Germany

DM 44,- / US $ 12,- / £ 5,- / NF 65,-
DM 12,- / US $ 3.25 / 1.40 / NF 17.50

6900 Heidelberg/Germany, Gaisbergstraße 6–8
P. O. Box 629
Cable Address: Groos Heidelberg

Banking Accounts:
No. 02/52338 Deutsche Bank Heidelberg
No. 1935936 Commerzbank Heidelberg

Printed by Julius Beltz, Weinheim/Germany

Published quarterly
Erscheint vierteljährlich

A TEST OF SPOKEN ENGLISH

H. Baetens Beardsmore and A. Renkin

L'article concerne une épreuve de production orale du type du dialogue, destinée à évaluer la qualité de la communication qu'un individu est capable d'établir dans une langue étrangère, en l'occurence, l'anglais. Lors de la construction de l'épreuve, les efforts ont porté tant sur le choix de son contenu linguistique que sur les aspects tels que la standardisation, l'objectivité et la rapidité de la correction. L'épreuve a été appliquée à 180 étudiants francophones du niveau universitaire – qui n'étaient pas de futurs professeurs d'anglais – qui devaient faire la preuve de leur connaissance pratique, active, de la langue. Les resultats d'une analyse d'items montrent que l'épreuve est de difficulté modérée et qu'elle possède un bon pouvoir de discrimination. Sa fidélité, calculée par la méthode du Split-half, atteint .86; elle est considérée comme satisfaisante.

La dernière partie de l'article est consacrée à une autre épreuve d'expression orale; cette épreuve complémentaire se trouve encore à l'état purement expérimental.

Der Artikel behandelt eine mündliche Sprachprüfung, die in Form eines Gesprächs durchgeführt wird und dazu dient, die Qualität der mündlichen Kommunikation in einer Fremdsprache, in diesem Falle Englisch, zu messen. Während der Erstellung des Testmaterials wurde nicht nur auf die Wahl des sprachlichen Inhalts geachtet, sondern auch auf solche Gesichtspunkte wie Standardisierung, Objektivität und Möglichkeit einer schnellen Korrektur. Dem Test haben sich 180 französischsprechende Universitätsstudenten – alle keine künftigen Englischlehrer – unterzogen, um ihre aktive Kenntnis der englischen Sprache unter Beweis zu stellen. Die Ergebnisse der Aufgabenanalyse beweisen, daß der Test einen mäßigen Schwierigkeitsgrad und eine gute Trennschärfe besitzt. Die Zuverlässigkeit des Tests, die unter Anwendung der 'Split-half'-Methode berechnet wurde, betrug 0,86. Sie darf als zufriedenstellend betrachtet werden.

Der letzte Teil des Artikels ist einem anderen Test der mündlichen Ausdrucksfähigkeit gewidmet, der sich aber noch in einem Experimentierstadium befindet.

I. *INTRODUCTION*

One of the most difficult fields in language testing has proved to be that of the productive spoken language. This is witnessed by the fact that so little has been done to date in this area.[1]

In the summary of decisions taken in May 1961 at the "Conference on Testing Proficiency of Foreign Students"[2] the necessity was expressed of urgently tackling research on techniques for measuring oral production. Eight years later, at the Second International Congress of Applied Linguistics (Cambridge, 1969) the need was again

1 Cf. Nehls (7), pp. 135–138.

2 Conference organised in Washington, in collaboration with other bodies, by the Center for Applied Linguistics of the Modern Language Association of America.

expressed for some means of testing the spoken language in a standardized manner which would produce information similar to that which could be obtained by an interview technique, while at the same time being more objective and less time-consuming.

At the Institute of Phonetics of the University of Brussels, where audio-oral language courses are given to students from the Faculty of Economics and the Business School, the problem of testing oral proficiency has been one of the foremost considerations. A test of spoken English has been designed which, it is hoped, is of a more uniform nature than other techniques, is easy and objective to score, is not time-consuming, and yet can provide adequate feed-back information to discriminate fairly between strong and weak students. The test forms part of a battery of tests used on pre-entry to a 180 hour intensive first-year preliminary course in spoken English. The whole group of tests enables one to classify students according to their level of proficiency in English and consequently to decide whether a strong student has a knowledge of English at least equal to that which he would gain by attending the course, thus allowing him to be dispensed.

Given the scope of the language teaching undertaken at the Institute of Phonetics, directed not towards language specialists, but towards students from science faculties requiring a practical working knowledge of English, it was decided to investigate techniques capable of providing information as to the quality of the active use of language, i.e. those techniques providing adequate information about oral production considered as a whole.[3]

Consequently it was decided not to use techniques which test one specific aspect only of language production, as in the case of lexical or supra-segmental features. There is no question of minimising the value of testing specific features of oral production, but as has been pointed out by Carroll (1), the choice of one's approach must depend on the aims pursued. Moreover, as has been noted by Grayshon (4), there is no clear-cut relationship between performance at tests of specific elements of language production and performance in a real-life situation. Thus, as has already been intimated, the spoken English test in use at the Institute of Phonetics was not designed to give a precise diagnostic in a limited domain, but was intended to give a

3 At the Institute of Phonetics other aspects of language have, of course, also been treated (8).

global appreciation of the
situation. It is intelligibility
said, that is the most impor

With this in mind, mir
and intonation could be s
the meaning of the commun

II. "GLOBAL" TECHNI(

1. The "strip-cartoon" tec

One of the most widely us
suggested by such authors
which consists in requestir
supposedly representing a
"strip-cartoon" technique
method of measuring ability
main defects; 1. the open-e
variety of responses; 2. the
language they may have lea
lary and constructions use
when it is evident that the
more extensive; 3. the fact
the type of language provid
not correlate with that supp
experiment directed by the
"strip-cartoon" technique
University students and to
having had a basic course i
the attitudes and interpret
the people depicted while th
factual and detailed descript
Thus it is easy to underst
difficult to correct and to giv
capacity for oral production.

2. The interview technique

Another widely used and
testing oral production is
strip-cartoon technique, in

ing the spoken language in a stand-
oduce information similar to that
erview technique, while at the same
time-consuming.
the University of Brussels, where
given to students from the Faculty
school, the problem of testing oral
foremost considerations. A test of
d which, it is hoped, is of a more
ues, is easy and objective to score,
n provide adequate feed-back infor-
een strong and weak students. The
sts used on pre-entry to a 180 hour
urse in spoken English. The whole
ssify students according to their
d consequently to decide whether a
English at least equal to that which
course, thus allowing him to be dis-

e teaching undertaken at the Institute
s language specialists, but towards
quiring a practical working know-
to investigate techniques capable of
quality of the active use of language,
dequate information about oral pro-

not to use techniques which test one
production, as in the case of lexical
re is no question of minimising the
of oral production, but as has been
oice of one's approach must depend
as has been noted by Grayshon (4),
between performance at tests of
duction and performance in a real-
been intimated, the spoken English
onetics was not designed to give a
domain, but was intended to give a

pects of language have, of course, also been

global appreciation of the quality of the communication in a dialogue situation. It is intelligibility, that is, the significance of what has been said, that is the most important criterion (Perren 8).

With this in mind, minor defects of pronunciation, accentuation and intonation could be safely ignored in so far as they do not alter the meaning of the communication or render it unintelligible.

II. *"GLOBAL" TECHNIQUES AVAILABLE*

1. The "strip-cartoon" technique

One of the most widely used techniques for testing oral production suggested by such authors as Lado (6), Valette (10) and others is that which consists in requesting a commentary upon a series of images supposedly representing a connected story – a development of the "strip-cartoon" technique used in essay-writing tests. However this method of measuring ability in the spoken language has *inter alia* three main defects: 1. the open-ended nature of such a test leads to a great variety of responses; 2. the students are in no way obliged to use what language they may have learned, and results have shown that vocabulary and constructions used tend to be rather limited and repetitive when it is evident that the student's knowledge of the language is more extensive; 3. the fact – and this is even more interesting – that the type of language provided by native speakers on such a test does not correlate with that supplied by foreign language learners. In an experiment directed by the Institute of Phonetics (not published) a "strip-cartoon" technique was administered to a group of British University students and to a group of Belgian University students having had a basic course in English; the former tended to describe the attitudes and interpret the emotions expressed in the features of the people depicted while the foreign-language learners tended to give factual and detailed descriptions of what they could see in the picture. Thus it is easy to understand why such tests are considered to be difficult to correct and to give an unsatisfactory picture of the student's capacity for oral production.

2. The interview technique

Another widely used and yet generally unsatisfactory technique for testing oral production is that of the interview. As opposed to the strip-cartoon technique, interviewing does have a few advantages,

particularly if the interview has been carefully prepared with several examiners judging each candidate's ability upon pre-established criteria. However, interviewing does have certain serious defects including the fact that it is time-consuming, wearisome for the examiner (and consequently unfair to students who have the misfortune of being questioned at the end of a lengthy session); interviews are difficult to standardize, almost impossible to score objectively, and difficult to keep independant from personality factors.

After having used the "strip-cartoon" technique for a considerable time at the Institute of Phonetics it was decided to abandon it in favour of interviews. These were to be administered for a provisional period which would allow the preparation and examination of a new technique capable of furnishing precise information in an accurate and standardized way, of cutting out subjective judgement and of facilitating administration and scoring.

3. The dialogue technique

The new test of spoken English in use for the past year is based upon the idea of a fictitious dialogue. The student is requested to react to certain stimuli in a particular way. The interlocutor is supposed to be perfect and provides stimuli which are realistic and grammatically correct and which under normal circumstances determine the type of response provided. In this way the student is led to participate in a dialogue with the voice presented on the tape-recorder so as to prove his ability to manipulate the spoken language in an active way that reflects his activity in an interview.

a) Collection and construction of the items

The items were carefully selected in order to leave very little scope for variations in the answers. This is of extreme importance if the test is to be standardized. The validity of the limited number of correct responses for each item was tested on a wide variety of native speakers, including very young children. Items were selected either on the basis of a simple contrastive analysis of the two languages involved or on what was considered basic to an elementary knowledge of spoken English. For our purposes it was decided that a selection had to be made from all the possible constructions one could expect from a student who had followed 180 hours of tuition in the spoken language, but this selection was not tied to the contents of any specific course.

It was found that in the co
to be avoided which would b
the general knowledge of the s
nature had to be of a kind wh
answer and any item likely to
hence inter-linguistic) clashes
such as "Where do your paren
later abandoned) brought out
through a wide variety of comp
lated, sometimes untranslata
linguistic clashes mentioned) to
All the stimuli had to be wi
leaving no room for ambiguity

b) Instructions

Before each section of items,
tongue (in our case Dutch
required from the student. Th
lish, of what the student resp
construction was required th
peated. The first item in the
scored and was intended to fam

c) Examples of contents

The basic abilities tested were
examples are not intended to
merely serve as illustrations): –

i) The manipulation of numb
S: Tell me when you wer
R: I was born on the fifth

ii) The correct use of adverbs
S: What do you usually d
R: I usually/normally dri

iii) The ability to provide sim
S: How did you come he
R: I came by tram/on foo

iv) The ability to agree with a
lar and irregular verbs
employed throughout the
S: I liked that novel, did
R: Yes, I did.

It was found that in the construction of the stimuli, questions had to be avoided which would bring into play the level of intelligence or the general knowledge of the students. Questions of a personal factual nature had to be of a kind which the student was willing and able to answer and any item likely to bring about violent cross-cultural (and hence inter-linguistic) clashes had to be avoided. For example, stimuli such as "Where do your parents live?" (used in the initial project and later abandoned) brought out answers varying from "I don't know" through a wide variety of complicated addresses sometimes half translated, sometimes untranslatable (the violent cross-cultural inter-linguistic clashes mentioned) to such things as "My parents are dead". All the stimuli had to be within the cultural grasp of the student leaving no room for ambiguity.

b) Instructions

Before each section of items, instructions were given in the native tongue (in our case Dutch or French) clearly indicating what was required from the student. This was followed by an example, in English, of what the student response should be like, and if a complex construction was required the instructions and example were repeated. The first item in the test was a "gift" item in that it was not scored and was intended to familiarize the students with the technique.

c) Examples of contents

The basic abilities tested were as follows (n.b. the items included as examples are not intended to give a complete summary of the test but merely serve as illustrations): –

i) The manipulation of numbers and dates, e.g.
   S: Tell me when you were born?
   R: I was born on the fifth of January, 1950.

ii) The correct use of adverbs, e.g.
   S: What do you usually drink at breakfast?
   R: I usually/normally drink tea at breakfast.

iii) The ability to provide simple personal factual information, e.g.
   S: How did you come here to day?
   R: I came by tram/on foot/in my father's car.

iv) The ability to agree with a speaker using the basic tenses of regular and irregular verbs (this usage of the verb was a criterion employed throughout the test), e.g.
   S: I liked that novel, did you?
   R: Yes, I did.

v) The ability to express disagreement with a speaker, e.g.
S: Will you be in on Sunday?
R: No, I won't.

vi) The ability to put a question to a speaker, e.g.
S: Ask me if I've been to Spain.
R: Have you been to Spain?

vii) The ability to indicate comprehension of a sequence of events or the relationship between cause and effect, e.g.
S: She bought a ticket, then she caught the train; what did she do before she caught the train?
R: She bought a ticket.

viii) The use of the simple past tense of irregular verbs and the natural use of personal pronouns in dialogue, e.g.
S: Did you meet my sister?
R: Yes, I met your sister yesterday.

ix) The correct use of the conditional, e.g.
S: She won't do any work because she's tired. And if she wasn't tired?
R: She'd do some work.

x) The correct use of the perfect conditional, e.g.
S: She wouldn't have seen him because he didn't arrive on time; and if he'd arrived on time?
R: She'd have seen him.

d) Correction procedure

As far as the correction of the test is concerned rigorous criteria must be applied at the outset. A strict principle of "right" or "wrong" on the face value of the response was adhered to, the student being treated as if he had to make himself perfectly intelligible to a native speaker – this avoided problems of interpreting whether the student had misunderstood the stimulus yet provided a correct answer or vice versa. Comprehension of spoken English was left to be tested more accurately by other means. Supra-segmental features of pronunciation, accentuation and intonation were not taken into account if the student had made himself understood accurately and clearly – on the other hand if he could not make himself understood because of wrong stress or bad pronunciation then he was not considered to have provided a fluent correct answer.

A printed correction sheet was used giving the most likely answers (including alternatives) and leaving room for variants to be written in–

correct variants would be inc
incorrect variants provided use

III. *RESULTS*

These results deal only with th
and (X) – 10 items – having
enough statistical information
lysis; preliminary information
however.

An analysis of the items ir
180 French-speaking subject
F.B. Davis (2). This techniqu
degree of difficulty and discri
be expressed as a mark varying

Table 1 shows the averag
culty and the discriminatory po

	D
average	50.
range	47

It is to be noted from this tab
there are no items which are
cult, while the average difficul
scale. Moreover, the 30 marks
helps to ensure the sensitivity
the whole range of marks, and
to be the threshold of success.

As far as the discriminatory
say that on the average it is re
that in effect any item with a
discriminatory power to be co
reasons we have thought that
grounds for being eliminated.
the distribution of the total sco
this distribution is normal and
deviation of 5.89.

correct variants would be incorporated into the marking sheet while incorrect variants provided useful information for error analysis.

III. *RESULTS*

These results deal only with the first 30 items in the test; sections (IX) and (X) – 10 items – having been added at a later date, there is not enough statistical information available yet to provide a detailed analysis; preliminary information does point in a favourable direction, however.

An analysis of the items in the test has been made on the results of 180 French-speaking subjects, using the technique suggested by F. B. Davis (2). This technique gives a simple means of estimating the degree of difficulty and discriminatory power of each item. These can be expressed as a mark varying from 0 to 100 on a linear scale.

Table 1 shows the average and the range of the degree of difficulty and the discriminatory power calculated on thirty items.

Table 1

	Difficulty	*Discrimination*
average	50.77	38.00
range	47=72–25	46=67–21

It is to be noted from this table that the test is of moderate difficulty; there are no items which are very easy and none which are very difficult, while the average difficulty is situated almost at the centre of the scale. Moreover, the 30 marks are well spread about the average; this helps to ensure the sensitivity of the measure revealed by the test over the whole range of marks, and particularly about the mean considered to be the threshold of success.

As far as the discriminatory power of the items is concerned we can say that on the average it is relatively high, if we consider with Davis, that in effect any item with a rating in excess of twenty has sufficient discriminatory power to be considered a useful item. For all the above reasons we have thought that none of the thirty items gave sufficient grounds for being eliminated, at least for the time being. Obviously the distribution of the total scores depends on the quality of the items; this distribution is normal and had an average of 14.96 and a standard deviation of 5.89.

Moreover a reliability coefficient has been calculated by the split-half method. The homogeneity of the averages and the variances on the two halves have been controlled. The correlation coefficient between the two halves, corrected by the Spearman-Brown formula gives a reliability coefficient of .86.

IV. *DISCUSSION*

One might be rather surprised to find that the analysis of the results has not led to the rejection or correction of certain items. However, a similar test had been previously tried out during a probationary period of a year and this preliminary test did provide substantial information as to the level of the questions, the quality of the instructions, etc., and enabled us to avoid repeating a certain number of both technical and content errors. It should be emphasized that with this sort of test even greater care than normal must be exercised in the choice of instructions since there is no means of recovering lost ground through erroneous interpretations of what the student was requested to do.

As far as the reliability of the test is concerned, it would seem to be very satisfactory, particulary since the test is never used in isolation, but always as part of a battery of five. Moreover, it is to be expected that this reliability will be even greater when the ten supplementary items added at a later date have been likewise analysed to the same extent as the thirty original items.

As yet we have not been able to estimate the validity of the test, given that there is no satisfactory external criterion available for this. However this problem should not be exaggerated since it is generally admitted (Guilford, 5, p. 388) that tests of ability have an a priori value if they have been constructed bearing in mind the well-identified aims of the teaching programme.

V. *THE COMPLEMENTARY TEST*

Although the tightly controlled productive test that we have just discussed provides ample information concerning the ability to manipulate certain constructions, it does differ basically from the interview technique in that it does not give the student an opportunity to prove his ability in connected flowing speech. In a search for a test of this ability to speak freely in a more natural manner experiments have been going on at the Institute of Phonetics to try and do this in a way which answers the same basic crite
lish test, namely criteria of
objectivity. To this end the
minute on a given topic. T
tape, to do something. (Ins
be asked to relate somethi
thoughts for a couple of mi
ments in advance. The mach
dent speaks for one minute.
title of the topic since the stu
arguments of a persuasive n
particular case the Dutch-sp
the listener to learn their la
invited to persuade the listen

For the correction of this
cribed and the number of e
test of oral production stric
weighting of the errors. To r
gested by Dodson (3) was a
measure of the relative value
with another.

Although this type of test
about a given topic, they ar
subject of their speech and
factory scale of comparison
is still in the experimental s
tailed analysis of its success.

VI. *CONCLUSIONS*

The information gathered fr
lish would seem to indicate
to measure the ability to ma
tain degree of accuracy, flu
measurement has been attain
comparison with other candi

4 The formula gives an order of
a percentage and is made up in

$$\left\{\frac{\text{n°. of words spoken}}{\text{n°. of errors made}} \times 10\right\}$$

answers the same basic criteria applied to the rest of the spoken English test. namely criteria of brevity, accuracy, ease of correction and objectivity To this end the students are requested to speak for one minute on a given topic. They are asked to persuade a listener, on tape, to do something. (Instead of persuading a listener they might be asked to relate something.) They are instructed to gather their thoughts for a couple of minutes so that they can prepare their arguments in advance. The machines are then set in motion and the student speaks for one minute. Care must be made in the selection of the title of the topic since the students must be willing and able to provide arguments of a persuasive nature on the subject concerned. In our particular case the Dutch-speaking students were invited to persuade the listener to learn their language while the French speakers were invited to persuade the listener to study in their particular faculty.

For the correction of this one-minute dialogue the tapes are transcribed and the number of errors noted. As with the more controlled test of oral production strict criteria must be applied concerning the weighting of the errors. To mark each transcription the formula[4] suggested by Dodson (3) was applied since this seems to give a good measure of the relative value of the student's efforts compared one with another.

Although this type of test does enable students to speak more freely about a given topic, they are still limited and standardized as to the subject of their speech and its length, and this provides a more satisfactory scale of comparison. However since this supplementary test is still in the experimental stage it is not yet possible to supply a detailed analysis of its success.

VI. *CONCLUSIONS*

The information gathered from the results of the test of spoken English would seem to indicate that it has achieved its purpose, this being to measure the ability to manipulate the spoken language with a certain degree of accuracy, fluency and intelligibility. Moreover, this measurement has been attained in a manner which allows for objective comparison with other candidates.

4 The formula gives an order of merit classification which can be transformed into a percentage and is made up in the following way: —

$$\left\{\frac{\text{n°. of words spoken}}{\text{n°. of errors made}} \times 10\right\} \qquad \left\{1 \pm \frac{\text{deviation from average n° of words}}{\text{fluency factor (500)}}\right\}$$

It overcomes many of the difficulties encountered with other techniques by its brevity, ease of application and correction – the whole test (excluding the one-minute speech discussed in section 5) takes 17 minutes to administer to a group of 24 students and by means of recording only the students' answers to the stimuli in a connected string with no blanks on the tape, each candidate's effort takes roughly 5 minutes to score. The complementary test (section 5) takes about 5 minutes to apply to 24 candidates and about 2 minutes per candidate to score. Basic requirements are tested in exactly the same way for all candidates, these requirements being considered (for our purposes) as the manipulation of the different forms of the verb in a varying structural environment.

Emphasis has been laid on the technical problems involved in constructing a standardized test of spoken English, since these have proved to hold many hidden snags which might appear absurd at first glance. Particular care had to be taken not only in the lay-out, instructions, timing, and so on, but also in the criteria used for scoring. Contents of the stimuli had likewise to be carefully vetted by native speakers so as to avoid ambiguous answers or too many variants (bringing scoring problems in their wake).

The supplementary test has been added to give students the opportunity to speak more freely using connected utterances while permitting the topic and time-factor involved to be controlled. The method of scoring used takes into account individual variations due to rate of speech and number of errors in proportion to the number of words spoken and enables a comparative study to be made of the test group.

A parallel version of the test of spoken English exists (using different lexical items) and it has been designed to furnish the same information as the version discussed in this paper. Although this version has not been tried out and analysed to the same extent as the first version it is presumed that it will be of the same general value since it was constructed with the same criteria in mind.

Our general conclusions would be that *at a given level of language ability* the information provided by the test of spoken English is sufficient to allow a replacement of the interview technique and that therefore it has achieved its purpose.

H. Baetens Beardsmore/A. Renkin
Institut de Phonétique
Université Libre de Bruxelles

BIBL

1. Carroll, J. B.: "The Psychology
Symposium. A Psycholinguistic App
University Press, 1968.
2. Davis, F. B.: *Analyse des items.* Louv
3. Dodson, J. C.: *Oral Examinations.*
Education, University of Wales, 19
4. Grayshon, M. C.: *The Examination*
mental weaknesses of examinatio
mother tongue is not English. N
tute of Education, 1968.
5. Guilford, J. P.: *Psychometric Method*
6. Lado, R.: *Language Testing. The C*
don: Longmans, Green and Co.
7. Nehls, D.: "Report of the 'Meet
Education' Held at the Residen
(Western Germany) From June 1
8. Perren, G. E.: "Testing Spoken L
guage Testing Symposium. A Psycho
Oxford University Press, 1968.
9. Renkin, A.: "L'évaluation des c
vivantes, 1970, sous presse.
10. Valette, R. M.: *Modern Language Te*
and World Inc., 1967.

ficulties encountered with other tech-
cation and correction – the whole test
discussed in section 5) takes 17 min-
24 students and by means of re-
s to the stimuli in a connected string
ch candidate's effort takes roughly
mentary test (section 5) takes about
tes and about 2 minutes per candi-
's are tested in exactly the same way
nents being considered (for our pur-
ne different forms of the verb in a

e technical problems involved in con-
spoken English, since these have
s which might appear absurd at first
taken not only in the lay-out, instruc-
so in the criteria used for scoring.
ewise to be carefully vetted by native
ious answers or too many variants
eir wake).

en added to give students the oppor-
z connected utterances while permit-
olved to be controlled. The method
nt individual variations due to rate of
proportion to the number of words
e study to be made of the test group.
spoken English exists (using differ-
designed to furnish the same infor-
in this paper. Although this version
ed to the same extent as the first ver-
e of the same general value since it
eria in mind.
ld be that *at a given level of language*
by the test of spoken English is
of the interview technique and that
e.

H. Baetens Beardsmore / A. Renkin
Institut de Phonétique
Université Libre de Bruxelles

BIBLIOGRAPHY

1. Carroll, J. B.: "The Psychology of Language Testing"; in: *Language Testing Symposium. A Psycholinguistic Approach* (Davies E., Editor). London: Oxford University Press, 1968.
2. Davis, F. B.: *Analyse des items.* Louvain: Nauwelaerts, 1966.
3. Dodson, J. C.: *Oral Examinations, Pamphlet n° 12.* Aberystwyth (Wales): Dept. of Education, University of Wales, 1963.
4. Grayshon, M. C.: *The Examination of Spoken English.* A discussion of some fundamental weaknesses of examinations in spoken English for students whose mother tongue is not English. Nottingham: University of Nottingham, Institute of Education, 1968.
5. Guilford, J. P.: *Psychometric Methods.* New York: Mc Graw Hill, 1964.
6. Lado, R.: *Language Testing. The Construction and Use of Foreign Language Tests.* London: Longmans, Green and Co. Ltd, 1961.
7. Nehls, D.: "Report of the 'Meeting of Experts on Language Testing in Adult Education' Held at the Residential College for Adult Education at Göhrde (Western Germany) From June 12 to 15, 1968", *IRAL* 7 (1969) pp. 133–148.
8. Perren, G. E.: "Testing Spoken Language: some unsolved problems"; in: *Language Testing Symposium. A Psycholinguistic Approach.* (Davies, E., Editor) London: Oxford University Press, 1968.
9. Renkin, A.: "L'évaluation des connaissances des langues", *Revue des langues vivantes,* 1970, sous presse.
10. Valette, R. M.: *Modern Language Testing. A Handbook.* New York: Harcourt, Brace and World Inc., 1967.